EP0881592

Publication Title:

Text/image selection from document images

Abstract:

Abstract of EP0881592

A method carried out in an image processing system in which images of documents are captured by an image capture device, such as a video camera, comprising: (a) displaying successive images captured by the video camera, each image being defined by greyscale image data and containing text matter, (b) receiving a first user input (mouse button click) defining the start of a selection and a first position within the displayed image, (c) in response to the first user input, freezing the displayed image, (d) determining the skew angle of text matter with respect to the field of view of the video camera, (e) receiving at least one further user input (further button click; drag of cursor), including a final user input (mouse button release), defining the end of a selection, and for the or each further user input, (f) determining, using the skew angle determined in step (d), the position, shape and dimensions of a selection element in dependence upon at least said first position, and (g) displaying the selection element superimposed on said frozen displayed image. The selection element may be a rectangle, or a selection vlock highlighting one or more words of text. Data supplied from the esp@cenet database - Worldwide d7c

---

Courtesy of http://v3.espacenet.com

This Patent PDF Generated by Patent Fetcher(TM), a service of Stroke of Color, Inc.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 881 592 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
02.12.1998 Bulletin 1998/49

(51) Int Cl.[6]: **G06K 9/32**, G06T 1/00

(21) Application number: **98304196.3**

(22) Date of filing: **28.05.1998**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **28.05.1997 GB 9711022**

(71) Applicant: **XEROX CORPORATION**
Rochester, New York 14644 (US)

(72) Inventors:
- **Taylor, Michael J., Dr.**
  61 Regent Street, Cambridge CB1 2B (GB)
- **Lee, Joe S.C.**
  Vancouver B.C. V5R 5R9 (CA)

(74) Representative: **Reynolds, Julian David et al**
Rank Xerox Ltd
Patent Department
Parkway
Marlow Buckinghamshire SL7 1YL (GB)

(54) Text/image selection from document images

(57) A method carried out in an image processing system in which images of documents are captured by an image capture device, such as a video camera, comprising: (a) displaying successive images captured by the video camera, each image being defined by greyscale image data and containing text matter, (b) receiving a first user input (mouse button click) defining the start of a selection and a first position within the displayed image, (c) in response to the first user input, freezing the displayed image, (d) determining the skew angle of text matter with respect to the field of view of the video camera, (e) receiving at least one further user input (further button click; drag of cursor), including a final user input (mouse button release), defining the end of a selection, and for the or each further user input, (f) determining, using the skew angle determined in step (d), the position, shape and dimensions of a selection element in dependence upon at least said first position, and (g) displaying the selection element superimposed on said frozen displayed image. The selection element may be a rectangle, or a selection vlock highlighting one or more words of text.

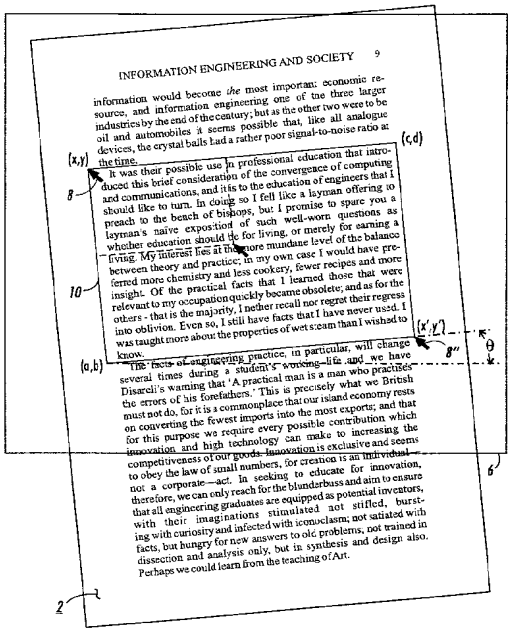


FIG. 2

EP 0 881 592 A2

Description

The present invention relates to image processing, and more particularly relates to techniques providing text/image selection from document images.

Conventional word processor applications for the personal computer enable a user to select text or image portions within a document, corresponding to an electronically stored file, by means of button presses and dragging of a mouse cursor.

The situation is quite different when the displayed document is that captured by a document camera providing greyscale, and usually relatively low resolution, images, such as those employed in over-the-desk scanning systems. It is known to use in such over-the-desk scanning systems a video camera disposed above a desk and capturing images of documents which are displayed to a user on a CRT monitor or other display device: these are discussed in detail, for example, in EP-A-622,722 (applicants' reference R/93003K/JDR) and British patent application 9614694.9 (applicants' reference R/96007/JDR). The capture of the document images may be for display in situ, or for transmission to a remote location as part of a videoconferencing tool.

However, a problem encountered with such systems is how to provide a very efficient text selection interface for interactive face-up document camera scanning applications. There is a need for techniques supporting the selection of rectangular regions of text and images within a captured by the camera via a "click-and-drag" of the mouse defining two points, or a leading diagonal, and for techniques providing, in much the same way as a word processor interface, for single and multi-word text selection from such a document.

The present invention provides a method carried out in an image processing system in which images of documents are captured by an image capture device, comprising: (a) displaying successive images captured by the image capture device, each image being defined by greyscale image data and containing text matter, (b) receiving a first user input defining the start of a selection and a first position within the displayed image, (c) in response to the first user input, freezing the displayed image, (d) determining the skew angle of text matter with respect to the field of view of the image capture device, (e) receiving at least one further user input, including a final user input defining the end of a selection, and for the or each further user input, (f) determining, using the skew angle determined in step (d), the position, shape and dimensions of a selection element in dependence upon at least said first position, and (g) displaying the selection element superimposed on said frozen displayed image.

The method preferably further comprises the step of: (h) extracting the image from within the selection element.

The method preferably further comprises the step of: (I) rotating the extracted image through the determined skew angle (θ), in the opposite sense.

The invention further provides a programmable image processing system when suitably programmed for carrying out the method of any of the appended claims or according to any of the particular embodiments described herein, the system including a processor, and a memory, an image capture device, an image display device and a user input device, the processor being coupled to the memory, image capture device, image display device and user input device, and being operable in conjunction therewith for executing instructions corresponding to the steps of said method(s).

In the case where the user employs a "click-and-drag" of the mouse defining two points, the remaining degree of freedom in the selection of a rectangle in the image is the skew of the text. The invention employs skew angle detection techniques to this document camera case where the location of "skew-pertinent" information is supplied by the user with the mouse, having extracted an intermediate image from the underlying greyscale image. The method is fast enough to find skew within less than 0.5s for most font sizes, which is fast enough to provide a pleasing interface. A similar effect is obtained for single-word and multi-word selection techniques.

Embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is view from above a desk of a document from which a text portion is to be selected in an over-the-desk scanning system according to an embodiment of the present invention;

Figure 2 shows the same view as in Fig. 1, after a user has finished selecting the text portion;

Figure 3 is a flow chart of the steps in providing text selection in accordance with an embodiment of the present invention;

Figure 4 shows the substeps employed in implementing the skew detection step in Fig. 3;

Figure 5 shows the effect of the substep in Fig. 4 of computing a high gradient image;

Figure 6 illustrates the effect of varying the size of the test image portion on the effect of the skew detection substep in Fig. 4;

Figure 7(a) shows a portion of captured and displayed text from which a user makes a selection, and Figure 7(b) shows in magnified form, part of the text matter of Fig. 7(a), showing a selected word;

Figure 8 is a flow chart showing the processing steps performed in providing the selection feedback illustrated in Fig. 7;

Figure 9 shows in more detail the technique (step s14 in Fig. 8) for determining the local inter-word threshold;
Figure 10 shows histogram is formed of horizontal gaps between the connected components of equal font size, (a) the ideal bimodal distribution, (b) real data with two attempted curve fittings, and (c) the fitting of a Gaussian curve;
Figure 11 illustrates text selection by diagonal sweep in an alternative embodiment of the invention; and
Figure 12 is a flow chart of the steps performed in providing the selection feedback shown in Fig. 11.

There are described below various techniques for text and/or image selection. It will be appreciated that these techniques may be used in conjunction with the image enhancement and thresholding techniques described in European patent application EP-A-_____ , based on British patent application 9711024.1 (applicants' ref: R/97008/JDR), filed concurrently herewith.

A. System configuration

It will be appreciated that the techniques according to the invention may be employed in any system or application where selection of a text portion from a multibit-per-pixel (e.g. greyscale or colour) image is required. Such instances include videoconferencing systems, scanning systems, especially the aforementioned over-the-desk scanning systems, multifunction devices, and the like. It will be appreciated that the invention may be implemented using a PC running Windows™, a Mac running MacOS, or a minicomputer running UNIX, which are well known in the art. For example, the PC hardware configuration is discussed in detail in *The Art of Electronics*, 2nd Edn, Ch. 10, P. Horowitz and W. Hill, Cambridge University Press, 1989. In the case of over-the-desk scanning, the invention may form part of the systems described in any of EP-A-495,622, EP-A-622,722, or European patent application EP-A-_____, based on British patent application 9614694.9 (applicants' reference R/96007/JDR) filed 12.7.96. The invention has been implemented in C++ on an IBM compatible PC running Windows® NT.

B. Rectangular text region selection via skew detection

This section describes a text selection technique that enables rectangular text region selection. The user defines a leading diagonal of the rectangle with a mouse. Automatic text skew detection is used to calculate the required image selection. Skew recovery is made efficient by analysing the image in the neighbourhood of the mouse input.

Figure 1 is view from above a desk of a document from which a text portion is to be selected in an over-the-desk scanning system incorporating an embodiment of the present invention.

Initially, a document 2 is open on the user's desk (not shown), and the user has positioned the document 2 so that the paragraph 4 which he wishes to scan/copy is within the field of view 6 of the camera (not shown). Images (greyscale) of the document 2 are captured and displayed to the user as feedback. As discussed in the EP-A-(R/96007/JDR), the content of the field 6 may be displayed (as live video images) within a window of any suitable display device, such as a CRT or LCD display. Using a conventional mouse, the user is able to control the cursor position in a familiar way; and the start of the selection of the paragraph 4 begins with the user pressing the left mouse button with the cursor at initial position 8. While the left mouse button remains pressed, the user makes a generally diagonal line (top left to bottom right): an intermediate cursor position 8' during this motion is shown.

Figure 2 shows the same view as in Fig. 1, after a user has finished selecting the text portion: end of selection by the user is inputted by the user releasing the left mouse button when the cursor is at the final cursor position 8". As can be seen, the text of document 2 is skewed with respect to the coordinate space of the camera's field of view 6: the angle of skew θ must be determined.

Figure 3 is a flow chart of the steps in providing text selection in accordance with an embodiment of the present invention. Initially, the start of selection user input is detected (step s1). Immediately (step s2), the image (i.e. within the field 6) displayed to the user is from on the display device (not shown). Next, a routine (s3) is performed to determine the skew angle θ, as is described in further detail below. Returning to Fig. 2, once the value of θ is obtained, the positions within the coordinate space of the display window of a selection rectangle 10 which is to be displayed as feedback to the user must be determined; the requirement being that, to provide a pleasing interface for the user, the selection rectangle 10 must be at the same skew angleθ. The coordinates ((x, y), (x', y')) corresponding to the initial and current, respectively, cursor positions 8, 8" are known. Using simple geometric relations, the coordinates (a, b) and (c, d) of the other corners of the rectangle 10 can readily be calculated. The skew angleθ is normally a small angle: generally it will be less than 5°.

As shown in Fig. 3, a rectangle is formed (step s5) with (x, y), (x', y'), (a, b) and (c, d) at the corners. This rectangle is then superimposed (step s6) on the stored frozen image data, and the resulting image displayed. At test is then made at step s7: if the user has finished selecting (i.e. an input received indicating that he has released the left mouse button), and if he has not, processing returns to step s4. (For illustration, the final cursor position 8" is used as the

'current' cursor position, although it will be appreciated that this process may be carried out continually during the user's diagonal movement of the cursor.)

If the user has finished selecting, the current image is frozen (step s8) in the display (window) Then, the image data for the image (here: the paragraph 4) present within the selection rectangle 10 is extracted (step s9) from that for the image within the field 6, and the extracted image is then rotated (step s10) through -θ, so as to ready it for further processing, such as OCR

Figure 4 shows the substeps employed in implementing the skew detection step in Fig. 3. This routine is based on the techniques described in US-A-5,187,753 and US-A-5,355,420 — maximising the variance of the laterally projected profile of differences over a range of skew angles, where the rotation of the image is made efficient by only performing vertical shears. The process begins (step s31) by initialising a grab area 12 (a small rectangular area to the right and below the cursor position such as that shown in Fig. 6 (discussed further below)). Suitably, the grab area 12 is just large enough for a few lines of text, and perhaps a couple of lines of 10 point text.

In order to minimise the amount of time taken to compute skew, we attempt to analyse the smallest amount of the image as we can. The algorithm is capable of resolving skew with less than two lines (it has been found) but the problem is clearly that it is not known how large the font size is before the skew angle has been determined.

To this end, an initial sample size (grab area 12) that is large enough to capture several lines of text at a "most likely" font size of between 10-12pt is used. Further, this initial region is to the right and below the initial "click", which assumes that the mouse is being dragged from top-left to bottom-right and that the skew angle is not too great (typically ±5 degrees). Because this is the most common font size that is selected using the interface, this gives an overall optimum response time.

The next step (s32) involves the computation of a high gradient image from the image within the initial grab area 12. The images of the document in Fig. 1 are greyscale images. An option is to threshold the image and then pass it to a skew detection algorithm. However, under uncontrolled lighting conditions, thresholding is potentially quite computationally expensive.

Figure 5 shows the effect of the substep in Fig. 4 of computing the high gradient image, which is accomplished using the familiar Sobel operator (discussed in more detail in Jähne B., *Digital Image Processing*, section 6.3.2, Springer-Verlag, 1991). In the resultant high gradient image of Fig. 5, each white pixel is the result of the gradient in the original (greyscale) image at that point being greater than a predetermined threshold, and each black pixel is the result of the gradient in the original (greyscale) image at that point being less than the predetermined threshold. The high-gradient image (used as the input for the skew detection) is easily computed from the greyscale supplied by the document camera and is a very reliable substitute for a fully thresholded image.

Computations are next performed on the image data for the high gradient image for each of the allowed set of skew angles (e.g. +5° to -5° in increments of 0.1°; although any suitable regime may be employed — see US-A-5,187,753 and US-A-5,355,420): step s33. In each case, the image is sheared (step s34) to approximate the rotation. Here, a technique used in the vertical shearing procedure that lies at the heart of the angular search is to wrap around the vertical shift. In other words, the pixels that are pushed out of the top of the region are re-inserted at the bottom in the same column. This way the variance profiles are always calculated on a rectangular region which makes everything neater, if not more reliable.

For the given angle, a lateral projection histogram for the image is computed (step s35). Based on the histogram data, the variance for the given angle is calculated (step s36). A plot of variance against angle (of rotation) may thus be plotted, as shown in Fig. 6(a). The ability of the technique to determine the skew angle depends on the size of the initial grab area 12 relative to the font size; and the absence of a discernible peak in Fig. 6(a) indicates that the computation has been unsuccessful. A test is made at step s37 to determine whether the highest peak in the plot of skew var *vs* angle is significant (a simple SNR based test), such as by determining whether the ratio of the peak value to the average value is greater than a predetermined value. If the peak is not significant, the size of the initial grab area 12 is increased (step s38), and the processing returns to step s32. The grab area 12 is expanded in the vertical direction more than the horizontal as it is in that direction that the most skew-pertinent information lies. This is done until an empirically defined threshold on the SNR (in this case defined to be the maximum variance divided by the mean variance) is reached.

Figures 6(b) and (c) illustrate the effect of varying the size of the grab area 12 on the effect of the skew detection substep in Fig. 4, in the case where the font size is 36pt. Clearly, a significant peak is ascertained for Fig. 6(b), from which a skew angle of 0.35° can be derived. This shows that very little text is needed for a good skew confidence: the grab area 12 of Fig. 6(b) is sufficient for the determination, and there is no need to expand to the larger area of Fig. 6 (c). In a preferred embodiment, the first grab area 12 is 100x100 pixels, the next largest is 200x200 pixels, and the next largest 300x300 pixels. If the latter fails, a value of θ=0 is returned.

The above description outlines the situation where text matter (paragraph 4) is sought to be selected by the user. However, the techniques according to the invention may be used for the selection of graphical objects within a document, and the aforementioned techniques have also been found to work well with graphics and line drawings.

The algorithm described in this section is very efficient, and the delay between starting to drag out the leading diagonal and the skew being detected is of the order of 0.5s on standard PC hardware, and slightly longer for larger and less common font sizes.

In addition, it will be appreciated that provision may be made, suitably using techniques for the resizing and moving of windows in the MS Windows environment, allowing the user to resize and/or reposition the selection rectangle 10 after it has been formed by releasing the left mouse button.

C. Single and multi-word selection methods

This section describes a single and multi-word text selection process, the aim being to imitate a common word processor interface, i.e. double click selects a word and "click-and-drag" may define a *non-rectangular* text region.

Figure 7(a) shows a portion 13 of captured and displayed text from which a user makes a selection. The selection may be of a single word, or of multiple consecutive words.

In this case, the user selects, using a cursor controlled by a mouse (not shown) in the conventional manner, from a portion 13 of displayed text matter a word 14 ("compression"): the user performs a "double-click" with the left mouse button with the mouse cursor in an initial position 8. As is shown (slightly exaggerated for the sake of illustration), the text matter is skewed by an angle θ with respect to the display coordinate space. Appropriate feedback must be displayed to the user, overlaid on the word 14, to show that it has been selected.

Figure 7(b) shows in magnified form, part of the text matter of Fig. 7(a), showing a selected word. To indicate selection, a selection block 20 is displayed overlaid on the word 14. (Here the block 20 is shown using hatching for the sake of illustration, but generally will comprise a solid black or coloured block, with the characters of the word 14 appearing as white or "reversed out".) The selection block 20 has vertical sides 22, 24 and horizontal sides 26, 28. The sides 22, 24 are positioned midway between the selected word 14 and the two adjacent words in the line— "analog" and "curve" respectively — and for this computations based on measured values of the inter-character separation (s_c) and the inter-word spacing (s_w) must be made, as described further below.

In addition, the sides 26, 28 are positioned midway between the line containing the selected word 14 and the line of text above and below it, respectively. The sides 26, 28 are also skewed byθ with respect to the horizontal dimension of the display, thereby providing appropriately-oriented selection feedback (block 20).

Figure 8 is a flow chart showing the processing steps performed in providing the selection feedback illustrated in Fig. 7. Initially (step s11), a user's double click of the left mouse button (i.e. first and second user inputs in rapid succession) is detected; and the displayed image is immediately frozen (step s12; although it will be appreciated that the freezing will occur upon the first of the "clicks" being made).

An operation is then performed (step s13), using a small region near (typically below and to the right of) the initial cursor position 8 (at the first mouse click), to determine the angle θ at which the text is skewed: this is described in detail above, with reference to Fig. 4. Then, a routine (s14) is performed to generate, for the small local region, an estimate of the inter-word spacing (threshold) $(s_w)_{min}$ (corresponding to $(s_c)_{max}$ — the threshold spacing above which the spacing must be an inter-word spacing rather than an inter-character spacing. A determination is then made (step s15), using known techniques, of the line separation within the small local region: this is the separation between the maximum height of characters on one line and the lowest level for characters on the line above it; and this enable the positions of the sides 26, 28 of the selection block 20 (Fig. 7(b)) to be determined.

In step s16 this determination is made, together with a calculation of the positions of the sides 22, 24 of the selection block 20: side 22 is $\frac{1}{2}(s_w)_{min}$ to the left of the character "c" in the selected word 14 ("compression"), and side 24 is $\frac{1}{2}(s_w)_{min}$ to the right of the "n" in the selected word 14. The selection block 20 with these sides is formed in step s17, and then in step s17 the selection block 20 is overlaid on the frozen image and the result displayed.

If the user has finished selecting, the current image is frozen (step s8) in the display (window). Then, the image data for the image (here: the word 14) present within the selection block 20 is extracted (step s19) from that for the image, and the extracted image is then rotated (step s20) through $-\theta$, so as to ready it for further processing, such as OCR. Figure 9 shows in more detail the technique (step s14 in Fig. 8) for determining the local inter-word threshold $(s_w)_{min}$. Initially (step s141), for the local region and using techniques known in the art for computing connected components, the character-character separations are measured for each pair of adjacent characters. Here, the previously obtained skew information (θ) is used to make the O'Gorman Docstrum techniques (Lawrence O'Gorman, "The Document Spectrum for Page Layout Analysis", in *IEEE Transactions On PAMI,* Vol 15, No. 11, Nov1993) run faster. O'Gorman used a connected component nearest neighbours method to find skew *and* inter-character and inter-line spacing. We use the skew information to find nearest neighbours in the line to give us inter-character information, and connected component heights to group blocks of consistent font size.

A histogram is formed of horizontal gaps between the connected components of equal font size. This is ideally is a bimodal distribution (see Fig. 10(a)), i.e. with a first peak (mode) 36 corresponding to inter-character spacings (s_c), and a second peak (mode) 38 corresponding to inter-word spacings (s_w). Figure 10(b) shows a plot 40 of the real

measured value for a typical sample, together with two curves sought to be fitted to the plot of the real data curve 40 — a curve 42 for a cubic fitting and a curve 44 for a quartic fitting. The curves 42, 44 intersect with the separation (distance) axis at I' and I", respectively.

The intersection (I') for the best fitting curve 42 may be used as the value of $(s_w)_{min}$. However, in the preferred embodiment, an attempt is made to make a best fit of a Gaussian curve to the first mode 36 in Fig. 10(a), and this is shown in Fig. 10(c). Obtaining the best fitting of the Gaussian curve 46 imputes an "intersection" I on the separation axis: this value is used as $(s_w)_{min}$ and must be determined.

Returning to Fig. 9, step s141 is followed by steps for finding the best fit, using 2k values about the value i=m, where m is the value at which the curve 46 has its peak. First, k is set to zero (step s142). In step s143, k is incremented by one, and then values for μ_k (the average of (2k+1) values around the mode) and σ_k^2 (the variance of (2k+1) values around the mode) are computed (step s144) according to the equations:

$$\mu_k = 1/(2k+1). \sum_{i=m-k}^{m+k} h(i) \qquad (1a)$$

$$\sigma_k^2 = 1/(2k+1). \sum_{i=m-k}^{m+k} \{h(i)- \mu_k\}^2 \qquad (1b)$$

The proportion of data in (2k+1) values around the mode is given by:

$$p_k = \{\sum_{i=m-k}^{m+k} h(i)\} / \{\sum_{I=0}^{N} h(i)\} \qquad (1c)$$

At step s145, an analysis is made of whether the curve is a good fit: this is done using the well known Chi-squared test. If the test is failed, processing returns to step s143. This is continued until the test of step s145 is passed. Upon the test being passed the value of $(s_w)_{min}$ has been found and is equal to $\mu_k+3\sigma_k$.

As can be seen, reliable determination of inter-character spacing enables us to segment out single words with a double click. Click-and-drag selection of multiple words needs further knowledge of the column limits. This is done using a technique due to Pavlidis (Pavlidis, T. and Zhou, J., "Page Segmentation and Classification," CVGIP, *Graphical Models and Image Processing*, Vol 54, No 6, Nov. 1992 pp. 484-496) based on vertical projection profiles at the connected component level. Again, the sample image used to establish column boundaries is grown until confidence measures are above specified levels. This interface has a slightly longer inherent delay than that of section B for finding skew alone, but with faster hardware this may become unimportant.

The technique of highlighting a word by a "double-click" has been described. It will be appreciated that, through simple modification of the techniques of Figs 8-10, and using well known techniques for finding sentence boundaries and the abovementioned methods of determining text (column) limits, techniques may be provided for indicating selection, in response to a third mouse click, of the whole sentence containing the word selected by the double click. I. e., the second click of the user's double click becomes an intermediate user input and the third click the final user input defining the end of selection.

Furthermore, it will be appreciated that, through simple modification of the techniques of Figs 8-10, and using the abovementioned methods of determining text (column) limits, techniques may be provided for indicating selection, in response to a selection comprising click, drag and release, of the whole of the text matter between the point of the initial click and the release of the mouse button. I.e., there is a first user input at the first mouse click (with the left mouse button being held down), an infinite number of intermediate "user inputs" as the cursor is dragged across the text, and the final user input defining the end of selection when the left mouse button is released. This is illustrated in Fig. 11 (the column limits are omitted for the sake of clarity/illustration).

Figure 12 is a flow chart of the steps performed in providing the selection feedback shown in Fig. 11. The process begins with the user's left mouse button click to start selection. The skew angle is determined as described hereinabove and then the column boundaries derived using techniques based on the abovementioned work of Pavlidis.

Next, a small vertical portion of the column is segmented. With knowledge of skew angle and column boundaries, this step could simply be to segment the whole column to locate the position of each of the words. However, word segmentation tends to be a slow operation, so instead, we divide the column into small horizontal strips and segment each of these separately. This segmentation process operates in a separate thread of program execution, thus allowing the user to freely continue moving the cursor, and for the system to update the selection display. This leads to a relatively fast interactive display, whilst at the same time, allowing anything from a single word, to a whole column to be selected.

As shown on Fig. 11, the selection block 30 thus formed is constituted by an upper block (covering the selected text on one line) extending from the left side 22' to a side coincident with the column boundary and having a lower side 32, and a lower block (covering the selected text on an adjacent line) extending from the side 24' and having an upper side 24. It will be appreciated that where the selection extends over 3 or more lines then the selection block also include an intermediate block, between the upper and lower ones, which extends between the column boundaries.

Copy to Clipboard

Once the user has selected a region, the next step is to copy it to the Windows *clipboard.* As previously described, this can be done in a number of different ways. The operations that are performed on the selected region prior to it being copied depend not only on the way in which it is copied, but also on the way in which it was selected; Table 1 highlights the operations necessary.

Table 1

Operations necessary for copying a selected region			
Selection method: *Copy as:*	**Rectangular selection box**	**Skewed selection box**	**Word-to-word selection**
Text	Copy region Binarise OCR	Copy region Rotate Binarise OCR	Copy region Rotate Mask unwanted text Binarise OCR
Color image	Copy region	Copy region Rotate	Copy region Rotate Mask unwanted text
Grey-scale image	Copy region Convert color to grey-scale	Copy region Rotate Convert color to grey-scale	Copy region Rotate Mask unwanted text Convert color to grey-scale
Binary image	Copy region Binarise	Copy region Rotate Binarise	Copy region Rotate Mask unwanted text Binarise

For example, if the region was selected using the skewed selection box and a color image is required, we first make a local copy of the selected region, de-skew by rotating it through the skew angle, and then place it on the clipboard. A more complex example is copying as text following a word-to-word selection. In this case, it is also necessary to mask out unwanted text from the beginning and end of the first and last lines of text. This is followed by converting the color image to a black and white image (binarisation), which is then passed to the OCR engine. Finally, the text returned by the OCR engine is then placed on the clipboard.

Of all these operations, one of the most important is the binarisation stage, particularly when followed by OCR. Due to the low resolution of the camera images, coupled with possible lighting variations, unacceptable results will be obtained if the camera image isbinarised using a simple threshold algorithm. Therefore, the image enhancement and thresholding techniques of British patent application 9711024.1 (ref. R/97008) are suitably used.

It will be further appreciated that a function may be provided (selectable by the user using a button on a toolbar of the UI) for when the user is invoking the 'Copy as text' function, enabling the line breaks to be removed from the OCRed text. This is useful, for example, when the text is to be pasted into a word processor. Furthermore, another such toolbar button may provide the option of the user viewing the selection they have copied in a preview window, in

a manner similar to the clipboard viewer on a conventional PC.

Claims

1. A method carried out in an image processing system in which images of documents are captured by an image capture device, comprising:

   (a) displaying successive images captured by the image capture device, each image being defined by grey-scale and/or colour image data and containing text matter,
   (b) receiving a first user input defining the start of a selection and a first position within the displayed image,
   (c) in response to the first user input, freezing the displayed image,
   (d) determining the skew angle (θ) of text matter with respect to the field of view of the image capture device,
   (e) receiving at least one further user input, including a final user input defining the end of a selection, and

   for the or each further user input,

   (f) determining, using the skew angle determined in step (d), the position, shape and dimensions of a selection element in dependence upon at least said first position, and
   (g) displaying the selection element superimposed on said frozen displayed image.

2. The method of claim 1, wherein step (d) comprises determining the skew angle of a first portion of text matter at or adjacent said first position.

3. The method of claim 1 or 2, wherein said final user input defines a second position within the displayed image, and said selection element comprises a rectangle having two opposing corners coincident with said first and second positions.

4. The method of claim 1 or 2, wherein the selection comprises the selection of one or more words of said text matter within said displayed image, and said selection element comprises a selection block overlaying said one or more words.

5. The method of claim 4, wherein step (f) comprises the substeps of:

   (f1) determining a word separation value $(s_w)_{min}$ from measured values of separation between adjacent pairs of characters in said text matter, and
   (f2) determining the dimensions of the selection block in the direction of flow of said text matter as a function of the word separation value $(s_w)_{min}$ determined in step (f1).

6. The method of claim 5, wherein step (f1) comprises the substeps of:

   (f1i) using a portion of said text matter, preferably at or adjacent said first position, forming a histogram of frequency versus inter-character spacing for each pair of adjacent characters within said portion,
   (f1ii) using a plurality of different Gaussian curves, determining which curve is a best fitting curve, said best-fitting curve forming a best fit with a predetermined mode of the histogram formed in step (f1i), and
   (f1iii) determining an estimate point on the inter-character spacing axis of the histogram at which the best-fitting curve satisfies a predetermined criteria.

7. The method of claim 6, wherein the estimate point corresponds to the value $(s_w)_{min} = \mu_k = +3\sigma_k$, where:

$$\mu_k = 1/(2k+1). \sum_{i=m-k}^{m+k} h(i) \quad ,$$

   and

$$\sigma_k^2 = 1/(2k+1). \sum_{i=m-k}^{m+k} \{h(i)- \mu_k\}^2.$$

8. The method of any of claims 4 to 7, wherein step (f) further comprises:

(f3) determining the line spacing (s_l) between adjacent lines of said text matter, and
(f4) determining the dimensions of the selection block in the direction perpendicular to the flow of said text matter as a function of said line spacing (s_l); and/or wherein step (f) further comprises:
(f5) determining the horizontal (column) limits of the text matter, and
(f6) determining whether said first and second positions are on different lines of said text matter.

9. The method of claim 9, wherein, if the determination in step (f6) is positive, step (f2) further comprises:

(f2i) for an upper portion of the selection block, overlaying text matter between said first position and the right hand horizontal (column) limit of the text matter, and
(f2ii) for a lower portion of the selection block, overlaying text matter between said second position and the left hand horizontal (column) limit of the text matter); and/or wherein step (f2) further comprises:
(f2iii) where the line of text containing said first position is separated by one or more further lines from the line of text containing said second position, for an internal portion of said selection block overlaying said one or more further lines, using as the left hand and right hand sides of said internal portion the left hand and right hand, respectively, horizontal (column) limits of said text matter.

10. A programmable image processing system when suitably programmed for carrying out the method of any of the preceding claims, the system including a processor, and a memory, an image capture device, an image display device and a user input device, the processor being coupled to the memory, image capture device, image display device and user input device, and being operable in conjunction therewith for executing instructions corresponding to the steps of said method(s).

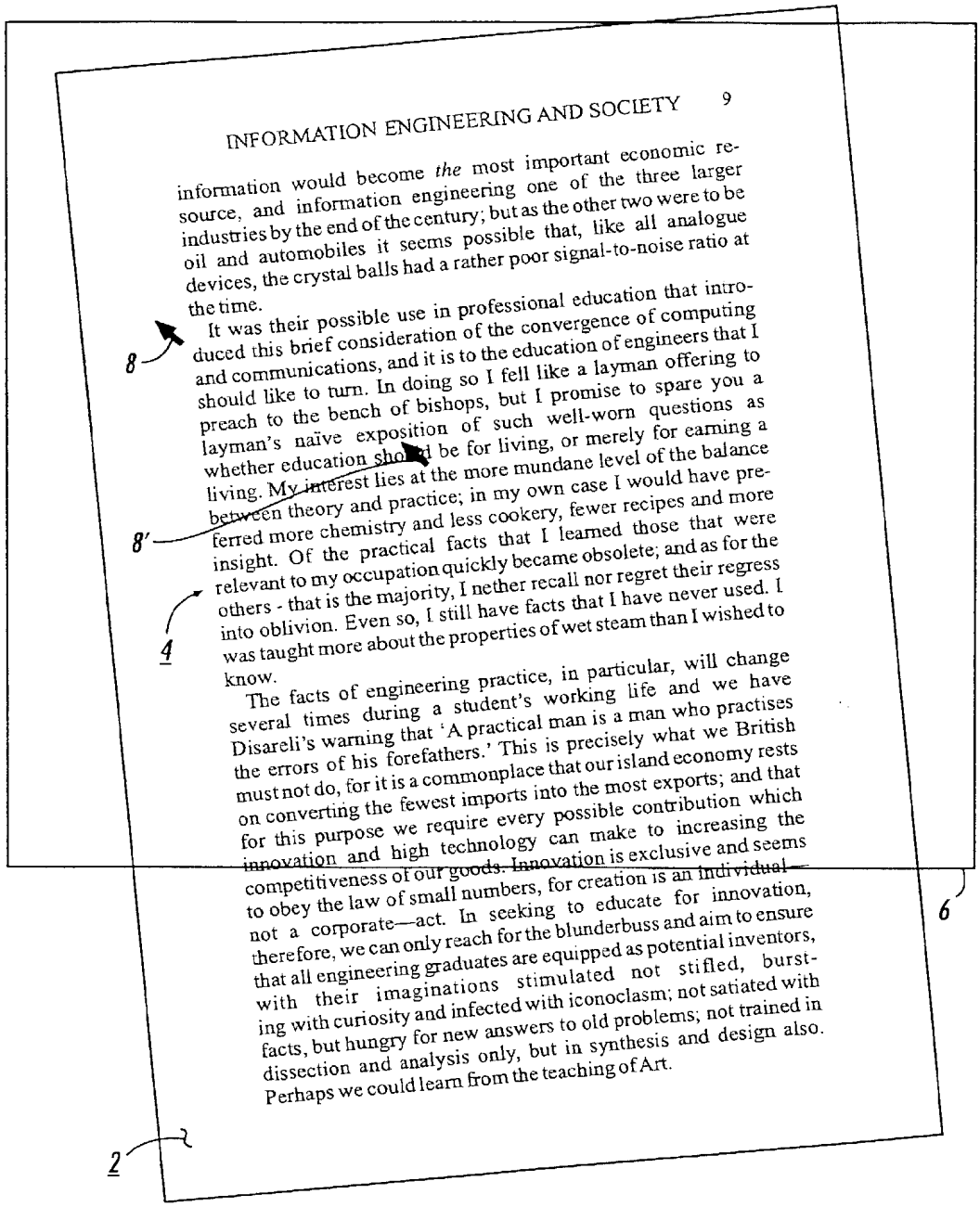

INFORMATION ENGINEERING AND SOCIETY 9
information would become *the* most important economic resource, and information engineering one of the three larger industries by the end of the century; but as the other two were to be oil and automobiles it seems possible that, like all analogue devices, the crystal balls had a rather poor signal-to-noise ratio at the time.
It was their possible use in professional education that introduced this brief consideration of the convergence of computing and communications, and it is to the education of engineers that I should like to turn. In doing so I fell like a layman offering to preach to the bench of bishops, but I promise to spare you a layman's naïve exposition of such well-worn questions as whether education should be for living, or merely for earning a living. My interest lies at the more mundane level of the balance between theory and practice; in my own case I would have preferred more chemistry and less cookery, fewer recipes and more insight. Of the practical facts that I learned those that were relevant to my occupation quickly became obsolete; and as for the others - that is the majority, I nether recall nor regret their regress into oblivion. Even so, I still have facts that I have never used. I was taught more about the properties of wet steam than I wished to know.
The facts of engineering practice, in particular, will change several times during a student's working life and we have Disareli's warning that 'A practical man is a man who practises the errors of his forefathers.' This is precisely what we British must not do, for it is a commonplace that our island economy rests on converting the fewest imports into the most exports; and that for this purpose we require every possible contribution which innovation and high technology can make to increasing the competitiveness of our goods. Innovation is exclusive and seems to obey the law of small numbers, for creation is an individual— not a corporate—act. In seeking to educate for innovation, therefore, we can only reach for the blunderbuss and aim to ensure that all engineering graduates are equipped as potential inventors, with their imaginations stimulated not stifled, bursting with curiosity and infected with iconoclasm; not satiated with facts, but hungry for new answers to old problems; not trained in dissection and analysis only, but in synthesis and design also. Perhaps we could learn from the teaching of Art.
8
8'
4
2
6

FIG. 1

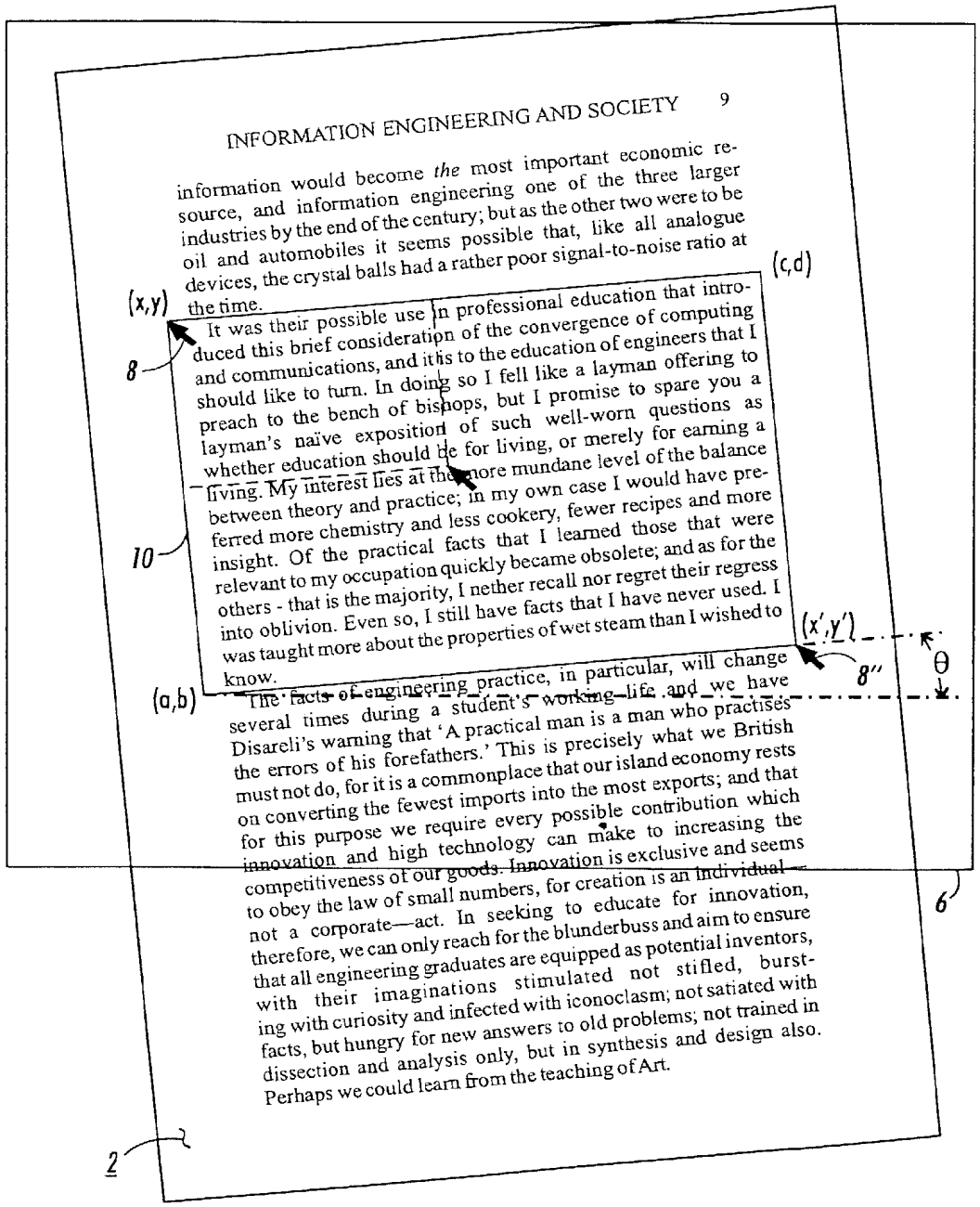

INFORMATION ENGINEERING AND SOCIETY 9
information would become *the* most important economic resource, and information engineering one of the three larger industries by the end of the century; but as the other two were to be oil and automobiles it seems possible that, like all analogue devices, the crystal balls had a rather poor signal-to-noise ratio at the time.
It was their possible use in professional education that introduced this brief consideration of the convergence of computing and communications, and it is to the education of engineers that I should like to turn. In doing so I fell like a layman offering to preach to the bench of bishops, but I promise to spare you a layman's naïve exposition of such well-worn questions as whether education should be for living, or merely for earning a living. My interest lies at the more mundane level of the balance between theory and practice; in my own case I would have preferred more chemistry and less cookery, fewer recipes and more insight. Of the practical facts that I learned those that were relevant to my occupation quickly became obsolete; and as for the others - that is the majority, I nether recall nor regret their regress into oblivion. Even so, I still have facts that I have never used. I was taught more about the properties of wet steam than I wished to know.
The facts of engineering practice, in particular, will change several times during a student's working life, and we have Disareli's warning that 'A practical man is a man who practises the errors of his forefathers.' This is precisely what we British must not do, for it is a commonplace that our island economy rests on converting the fewest imports into the most exports; and that for this purpose we require every possible contribution which innovation and high technology can make to increasing the competitiveness of our goods. Innovation is exclusive and seems to obey the law of small numbers, for creation is an individual — not a corporate—act. In seeking to educate for innovation, therefore, we can only reach for the blunderbuss and aim to ensure that all engineering graduates are equipped as potential inventors, with their imaginations stimulated not stifled, bursting with curiosity and infected with iconoclasm; not satiated with facts, but hungry for new answers to old problems; not trained in dissection and analysis only, but in synthesis and design also. Perhaps we could learn from the teaching of Art.
(x,y)
(c,d)
(a,b)
(x',y')
θ
8
8''
10
2
6

FIG. 2

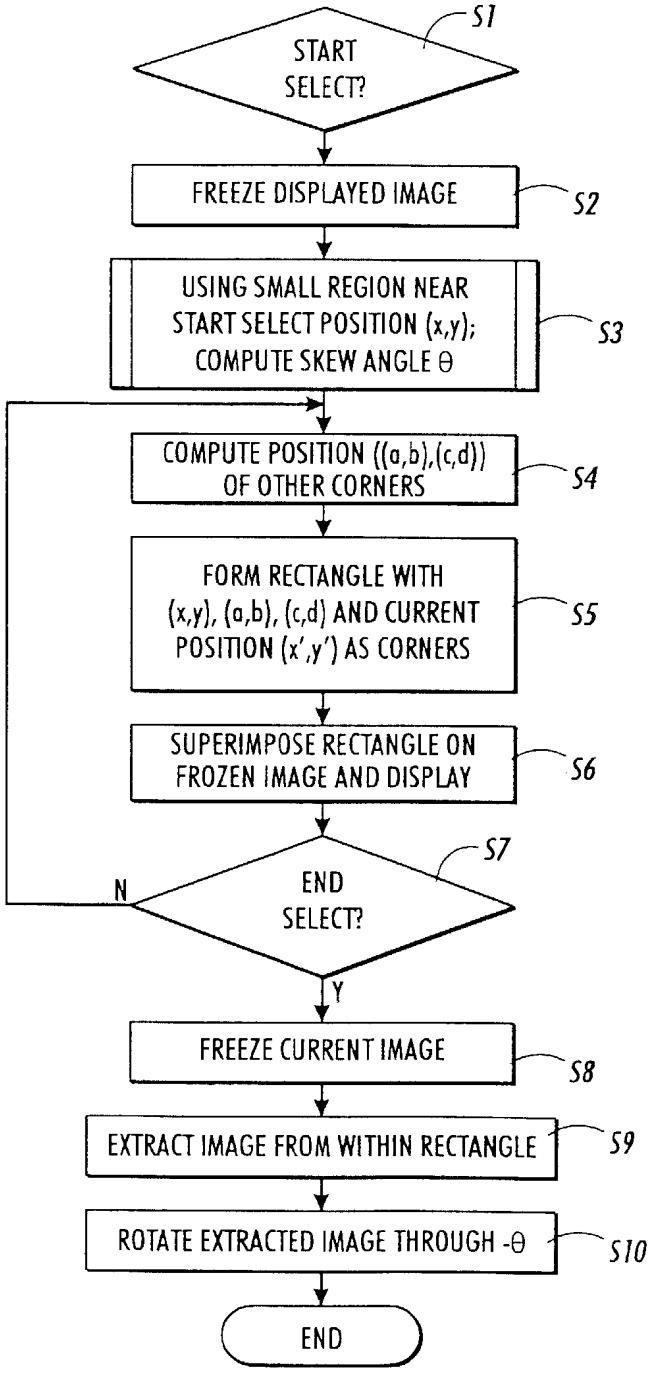
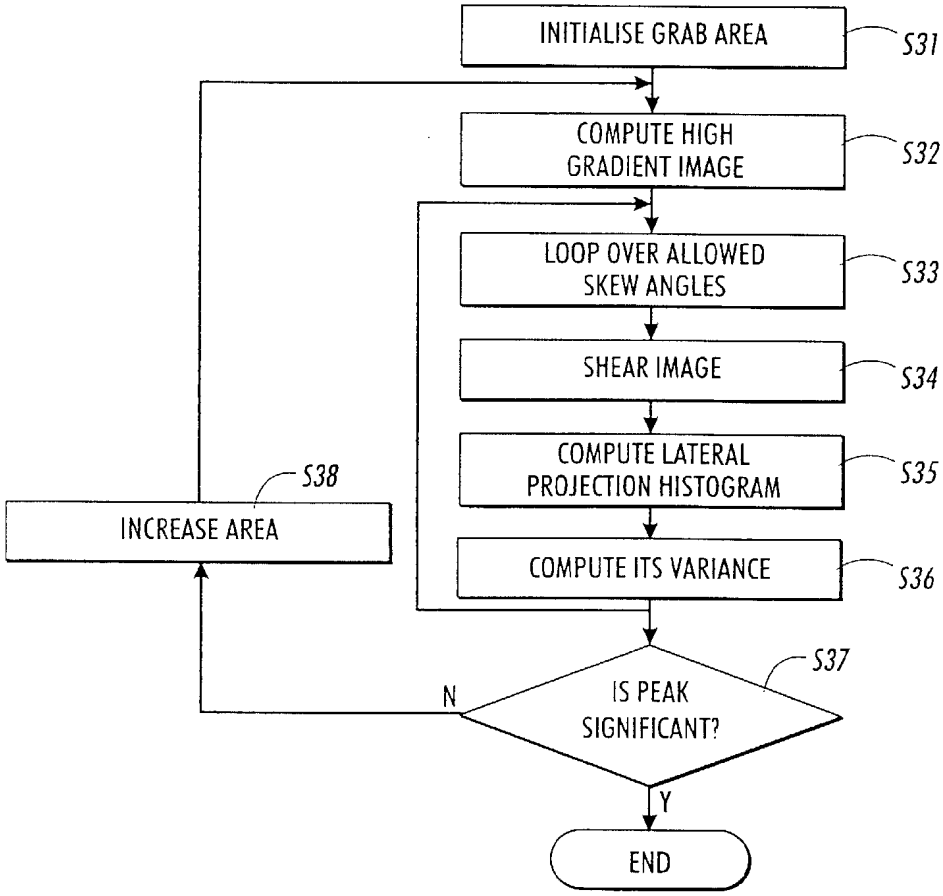

S1
START SELECT?
FREEZE DISPLAYED IMAGE
S2
USING SMALL REGION NEAR START SELECT POSITION (x,y); COMPUTE SKEW ANGLE θ
S3
COMPUTE POSITION ((a,b),(c,d)) OF OTHER CORNERS
S4
FORM RECTANGLE WITH (x,y), (a,b), (c,d) AND CURRENT POSITION (x',y') AS CORNERS
S5
SUPERIMPOSE RECTANGLE ON FROZEN IMAGE AND DISPLAY
S6
S7
END SELECT?
N
Y
FREEZE CURRENT IMAGE
S8
EXTRACT IMAGE FROM WITHIN RECTANGLE
S9
ROTATE EXTRACTED IMAGE THROUGH -θ
S10
END

FIG. 3

FIG. 4


INITIALISE GRAB AREA
S31
COMPUTE HIGH GRADIENT IMAGE
S32
LOOP OVER ALLOWED SKEW ANGLES
S33
SHEAR IMAGE
S34
COMPUTE LATERAL PROJECTION HISTOGRAM
S35
COMPUTE ITS VARIANCE
S36
IS PEAK SIGNIFICANT?
S37
N
Y
INCREASE AREA
S38
END

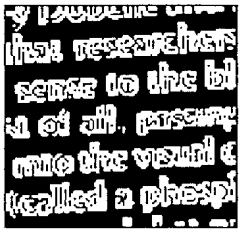


FIG. 5

FIG. 6A

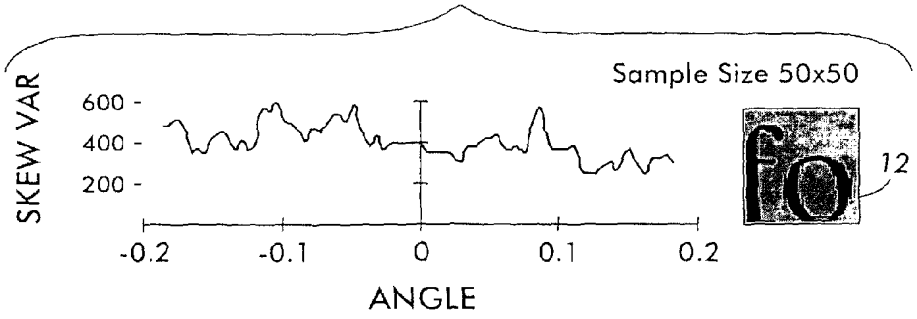
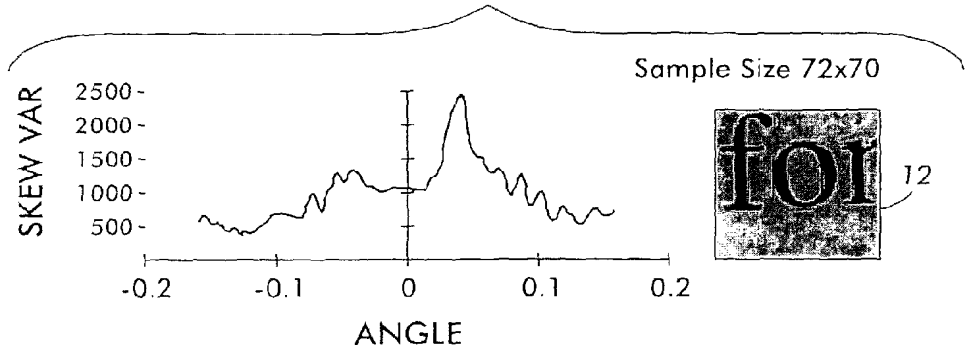


FIG. 6B



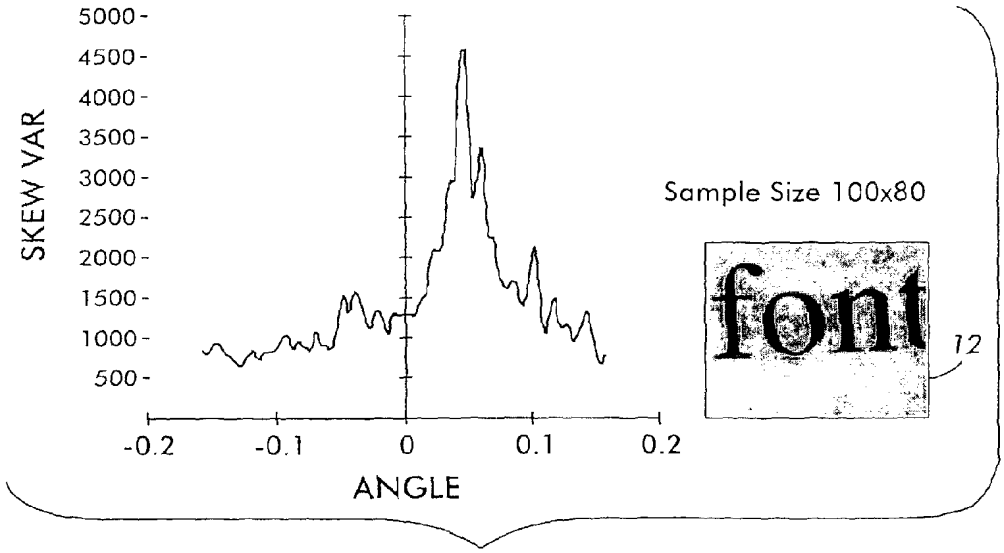
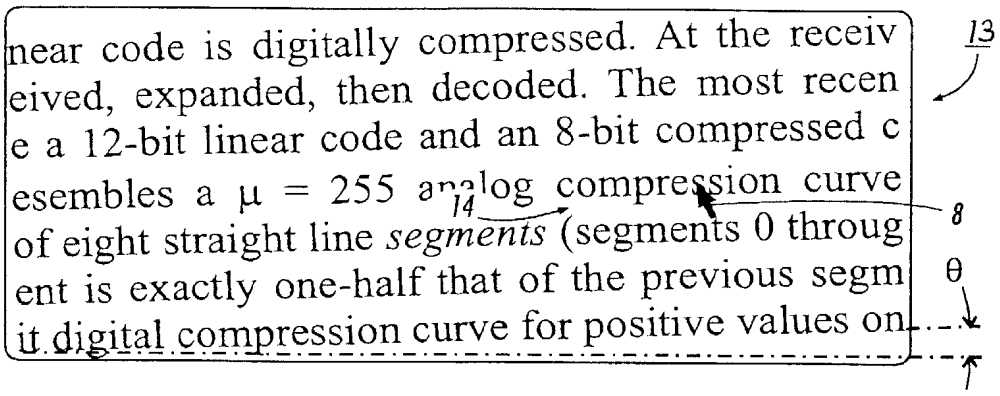


FIG. 6C

FIG. 7A


near code is digitally compressed. At the receiv
eived, expanded, then decoded. The most recen
e a 12-bit linear code and an 8-bit compressed c
esembles a μ = 255 analog compression curve
of eight straight line *segments* (segments 0 throug
ent is exactly one-half that of the previous segm
it digital compression curve for positive values on
13
14
8
θ

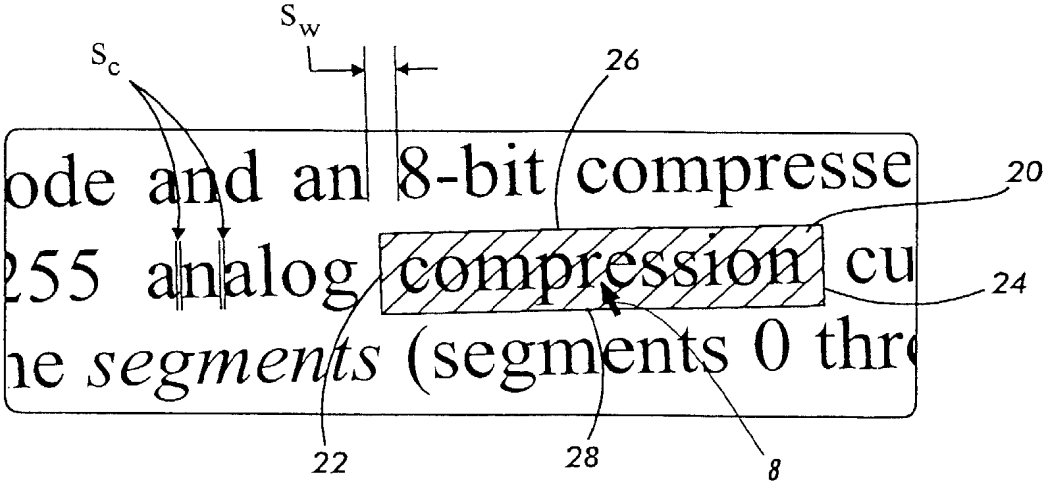

S_w
S_c
26
ode and an 8-bit compresse
255 analog compression cu
ne *segments* (segments 0 thr
20
24
22
28
8

FIG. 7B

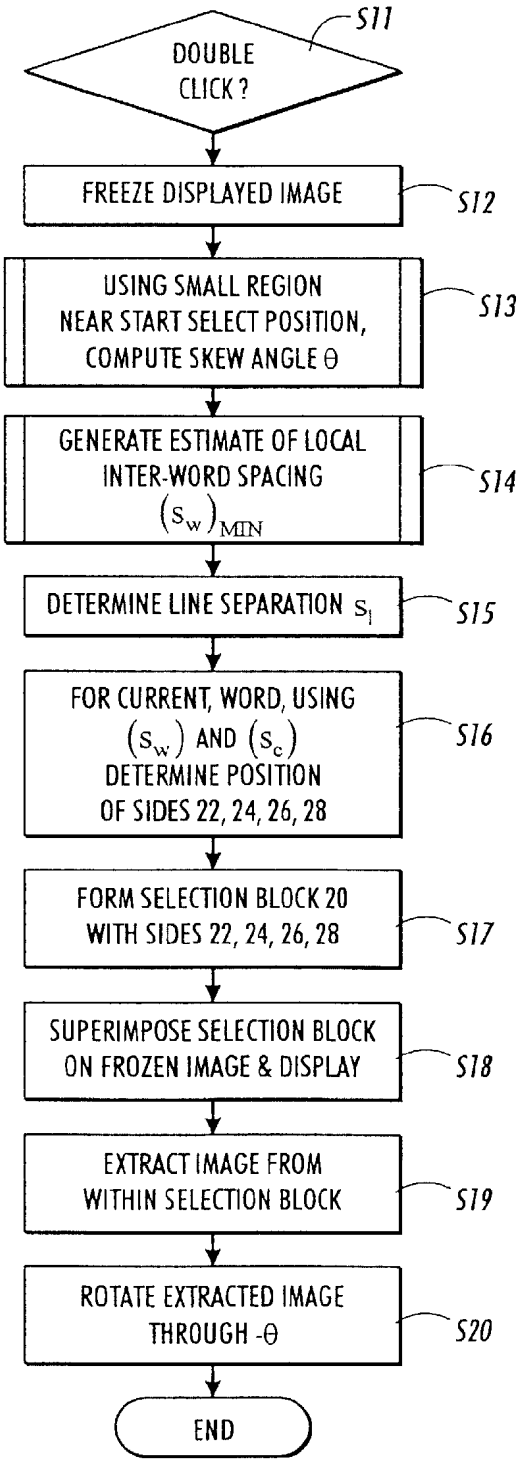

S11
DOUBLE CLICK ?
FREEZE DISPLAYED IMAGE
S12
USING SMALL REGION NEAR START SELECT POSITION, COMPUTE SKEW ANGLE θ
S13
GENERATE ESTIMATE OF LOCAL INTER-WORD SPACING $(s_w)_{MIN}$
S14
DETERMINE LINE SEPARATION s_l
S15
FOR CURRENT, WORD, USING (s_w) AND (s_c) DETERMINE POSITION OF SIDES 22, 24, 26, 28
S16
FORM SELECTION BLOCK 20 WITH SIDES 22, 24, 26, 28
S17
SUPERIMPOSE SELECTION BLOCK ON FROZEN IMAGE & DISPLAY
S18
EXTRACT IMAGE FROM WITHIN SELECTION BLOCK
S19
ROTATE EXTRACTED IMAGE THROUGH -θ
S20
END

FIG. 8

FIG. 9

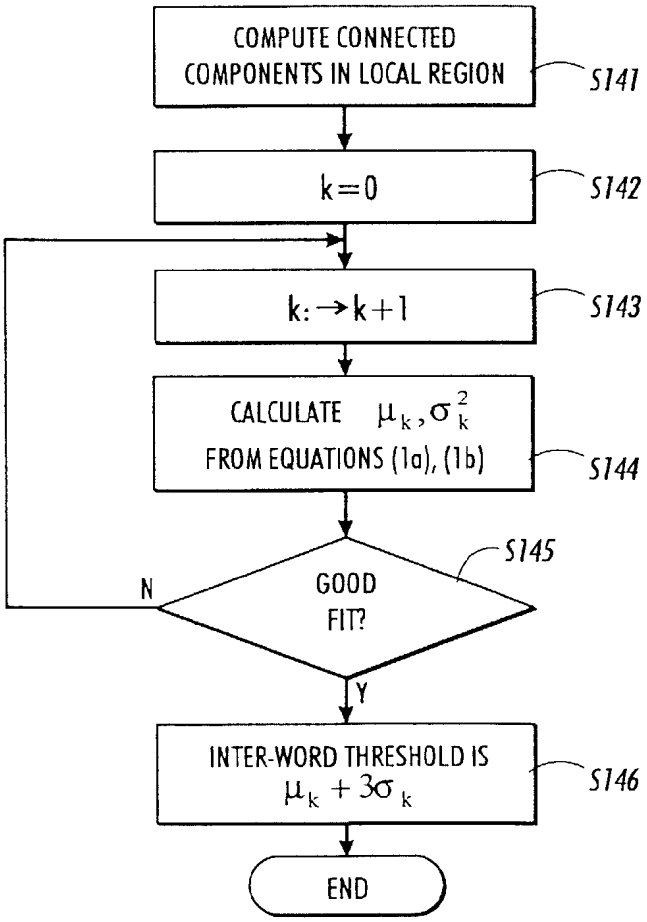
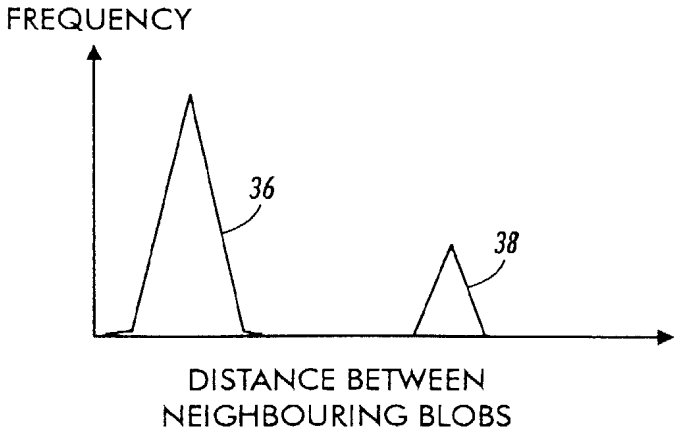

COMPUTE CONNECTED COMPONENTS IN LOCAL REGION
S141
k=0
S142
k:→k+1
S143
CALCULATE μ_k, σ_k^2 FROM EQUATIONS (1a), (1b)
S144
GOOD FIT?
S145
N
Y
INTER-WORD THRESHOLD IS $\mu_k + 3\sigma_k$
S146
END

FIG. 10A



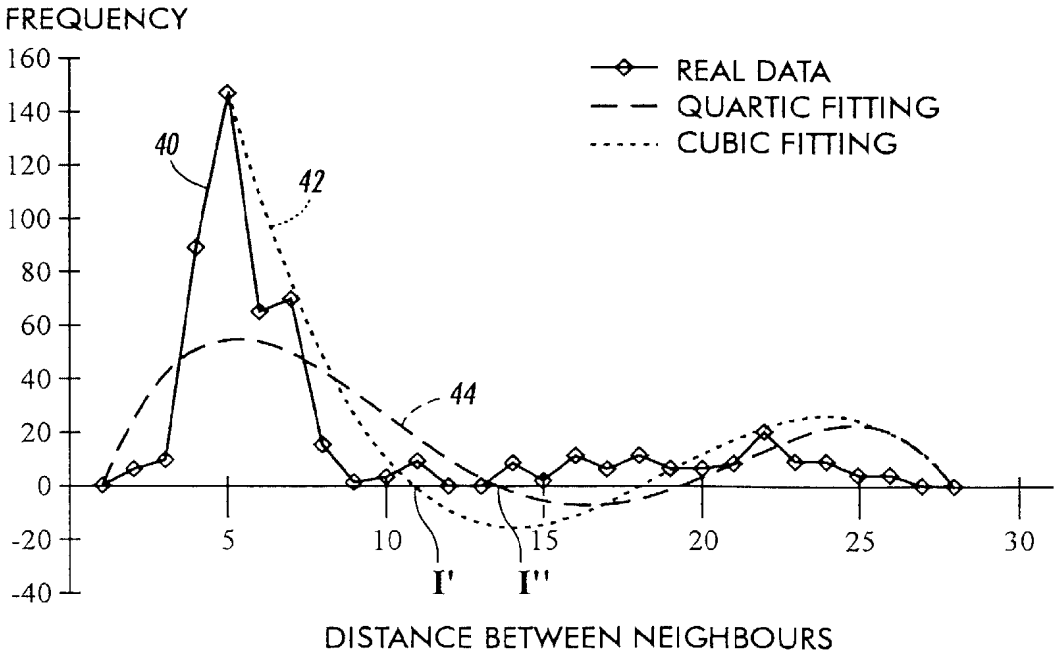


FIG. 10B

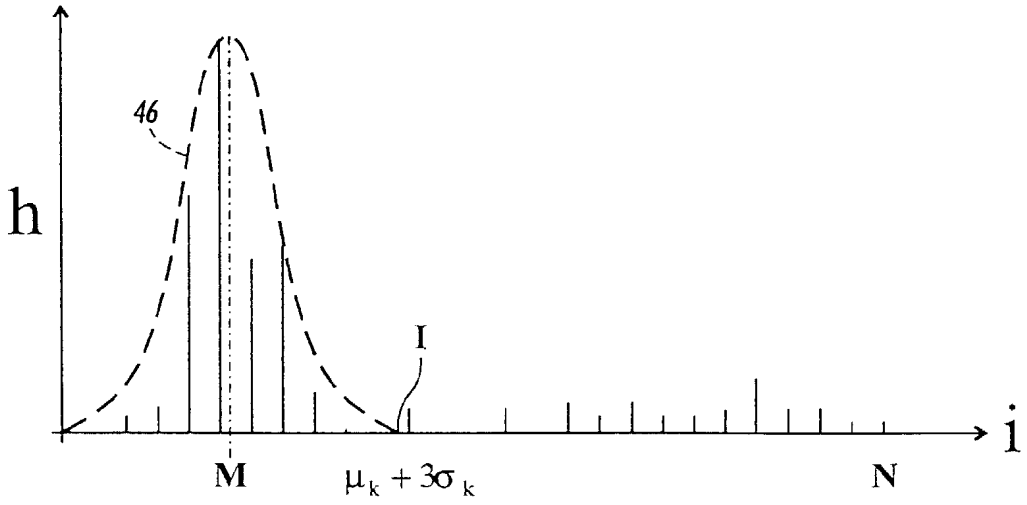


FIG. 10C

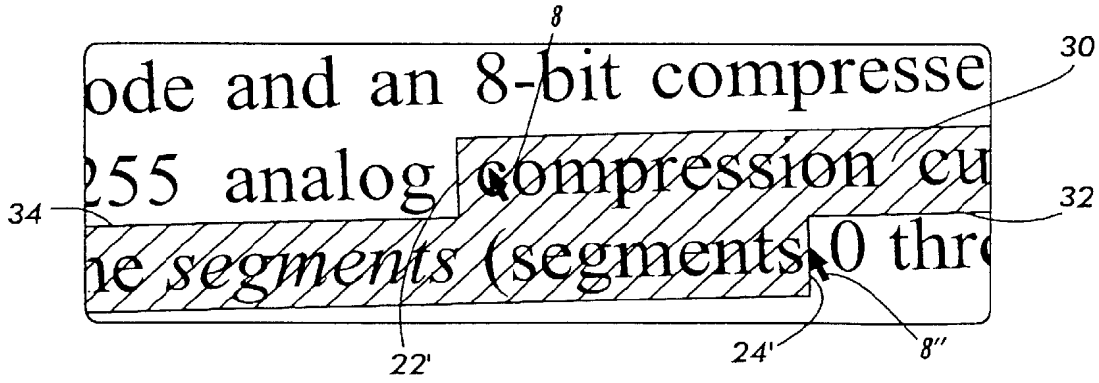

8
30
ode and an 8-bit compresse
255 analog compression cu
32
34
ne segments (segments 0 thr
22'
24'
8''

FIG. 11

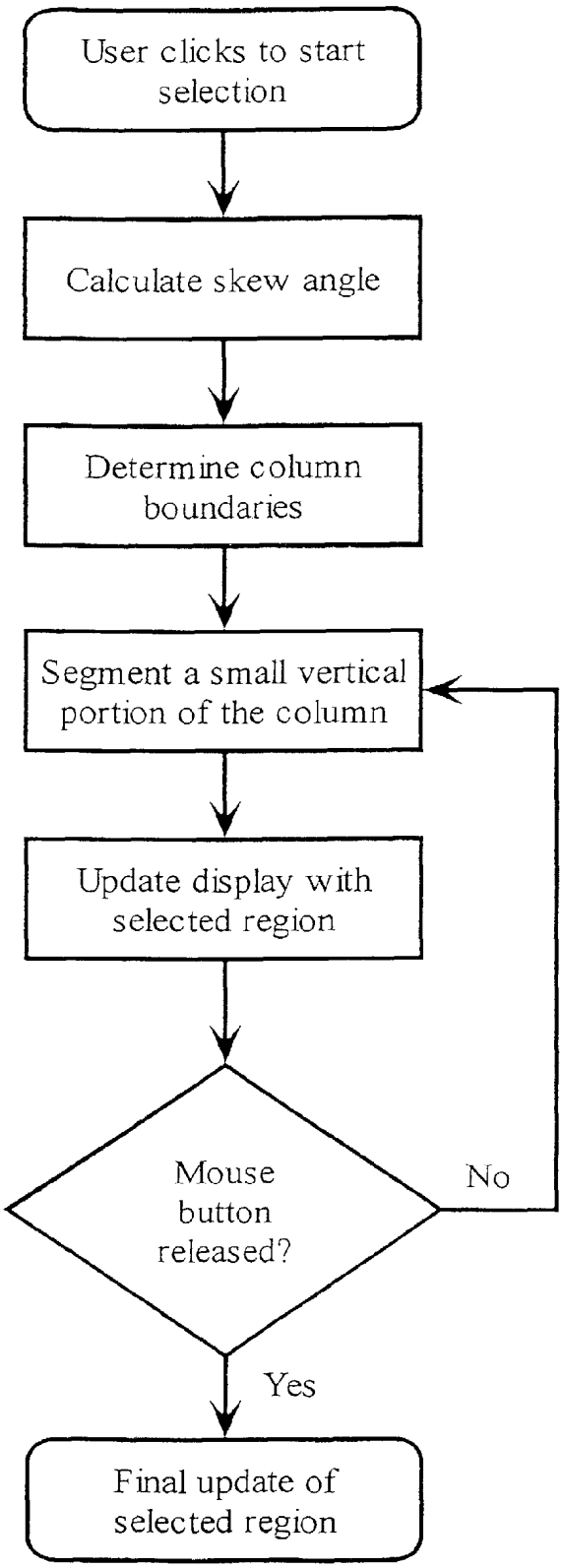

User clicks to start selection
Calculate skew angle
Determine column boundaries
Segment a small vertical portion of the column
Update display with selected region
Mouse button released?
No
Yes
Final update of selected region

Fig. 12